Exact bounds on the closeness between the Student and standard normal distributions

Iosif Pinelis*

Department of Mathematical Sciences
Michigan Technological University
Houghton, Michigan 49931, USA
E-mail: ipinelis@mtu.edu

Abstract: Upper bounds on the Kolmogorov distance (and, equivalently in this case, on the total variation distance) between the Student distribution with p degrees of freedom (SD_p) and the standard normal distribution are obtained. These bounds are in a certain sense best possible, and the corresponding relative errors are small even for moderate values of p. The same bounds hold on the closeness between SD_p and SD_q with $q > p$.

AMS 2000 subject classifications: Primary 62E17; secondary 60E15, 62E20, 62E15.
Keywords and phrases: Student's distribution, standard normal distribution, Kolmogorov distance, total variation distance, probability inequalities.

Contents

1. Summary and discussion

The density and distribution functions of Student's distribution with p degrees of freedom (SD_p) are given, respectively, by the formulas

$$f_p(x) := \frac{\Gamma\left(\frac{p+1}{2}\right)}{\sqrt{\pi p}\,\Gamma\left(\frac{p}{2}\right)} \left(1 + \frac{x^2}{p}\right)^{-(p+1)/2} \qquad \text{and} \tag{1.1}$$

$$F_p(x) := \int_{-\infty}^{x} f_p(u)\,\mathrm{d}u \tag{1.2}$$

for all real x. Most often, the values of the parameter p are assumed to be positive integers. However, formula (1.1) defines a probability density function

*Supported by NSF grant DMS-0805946

for all real $p > 0$, and, as we shall see, it may be advantageous, at least as far as proofs are concerned, to let p take on all positive real values. Let us also extend definitions (1.1) and (1.2) by continuity to $p = \infty$, so that

$$f_\infty =: \varphi \text{ and } F_\infty =: \Phi$$

are the density and distribution functions of the standard normal distribution (SND).

The standard normal and Student distributions are clearly among the most common distributions in statistics. It is a textbook fact that the SD_p is close to the SND when p is large, say in the sense that $f_p(x) \underset{p\to\infty}{\longrightarrow} f_\infty(x)$ for each real x. By Scheffé's theorem [8], this implies the convergence of the total variation distance

$$d_{\mathrm{TV}}(p) = \frac{1}{2}\int_{-\infty}^{\infty} |f_p(x) - \varphi(x)|\,\mathrm{d}x \tag{1.3}$$

to 0 as $p \to \infty$. In fact, the convergence of the SD_p to the SND is presented in [8] as the motivating case.

Consider also the Kolmogorov distance

$$d_{\mathrm{Ko}}(p) := \sup_{x\in\mathbb{R}} |F_p(x) - \Phi(x)|$$

between the SD_p and SND. It is clear that, for any two probability distributions, the Kolmogorov distance between them is no greater than twice the total variation distance, and hence the convergence of the latter distance to 0 implies that of the former.

However, in the present case one can say more. For any p and q in the interval $(0, \infty]$, let $d_{\mathrm{Ko}}(p, q)$ and $d_{\mathrm{TV}}(p, q)$ denote, respectively, the Kolmogorov distance and the total variation distance between SD_p and SD_q, so that $d_{\mathrm{Ko}}(p) = d_{\mathrm{Ko}}(p, \infty)$ and $d_{\mathrm{TV}}(p) = d_{\mathrm{TV}}(p, \infty)$.

Proposition 1.1.

(i) For all p and q such that $0 < p < q \leqslant \infty$

$$\tfrac{1}{2} d_{\mathrm{TV}}(p, q) = d_{\mathrm{Ko}}(p, q) = \max_{x\in(0,\infty)} \big(F_q(x) - F_p(x)\big). \tag{1.4}$$

(ii) Moreover, for each $p \in (0, \infty)$ the distance $d_{\mathrm{Ko}}(p, q)$ is strictly increasing in $q \in [p, \infty]$, and for each $q \in (0, \infty]$ the distance $d_{\mathrm{Ko}}(p, q)$ is strictly decreasing in $p \in (0, q]$. In particular,

$$d_{\mathrm{Ko}}(p, q) < d_{\mathrm{Ko}}(p, \infty) = d_{\mathrm{Ko}}(p) \tag{1.5}$$

for all p and q such that $0 < p \leqslant q < \infty$, and $d_{\mathrm{Ko}}(p)$ is strictly decreasing in $p \in (0, \infty]$.

Statement (ii) holds as well with d_{TV} in place of d_{Ko}.

This proposition and the other results stated in this section will be proved in Section 2.

The Kolmogorov distance and the total variation one are apparently the two most commonly used distances between probability distributions. Therefore, it seems natural to consider the rate of convergence of $d_{\mathrm{Ko}}(p)$ and, equivalently, $d_{\mathrm{TV}}(p)$ to 0 as $p \to \infty$, which is part of what is done in this paper. Actually, the motivation for this study comes from the discussion in [4]. In turn, the paper [4] was motivated by developments of [6].

Theorem 1.2. *For any real $p \geqslant 4$*

$$\tfrac{1}{2}\, d_{\mathrm{TV}}(p) = d_{\mathrm{Ko}}(p) < C/p, \tag{1.6}$$

where

$$C := \frac{1}{4}\sqrt{\frac{7+5\sqrt{2}}{\pi e^{1+\sqrt{2}}}} = 0.158\ldots. \tag{1.7}$$

Moreover,

$$\lim_{p\to\infty} p\, d_{\mathrm{Ko}}(p) = C, \tag{1.8}$$

so that the constant C is the best possible one in (1.6).

In what follows, it is assumed by default that

$$a := 1/p.$$

Theorem 1.2 is based on

Theorem 1.3. *For any real $p \geqslant \frac{50}{29}$*

$$\tfrac{1}{2}\, d_{\mathrm{TV}}(p) = d_{\mathrm{Ko}}(p) < B(a, \tilde{x}_a), \tag{1.9}$$

where

$$B(a,x) := \frac{a}{768}\Big(8x\big[5a^2x^2 + a(3x^6 - 7x^4 - 5x^2 - 3) + 24(x^2+1)\big]\varphi(x) \\ + 33a^2(2\Phi(x) - 1)\Big)$$

and $\tilde{x}_a$ is, for any $a \in (0,1)$, the unique real root $x > 0$ of the polynomial equation

$$P(a,x) := -96(x^4 - 2x^2 - 1) - 4a(3x^8 - 28x^6 + 30x^4 + 12x^2 + 3) \\ - a^2(20x^4 - 60x^2 - 33) = 0. \tag{1.10}$$

In fact, it will be shown (Lemma 2.6) that

$$B(a, \tilde{x}_a) < C/p \tag{1.11}$$

for all $p \geqslant 4$.

Note that, since the polynomial equation (1.10) is of degree 4 in x^2, the root $\tilde{x}_a$ can be expressed in radicals of polynomials in a.

By the triangle inequality, (1.6) implies that $\frac{1}{2}\, d_{\mathrm{TV}}(p,q) = d_{\mathrm{Ko}}(p,q) \leqslant d_{\mathrm{Ko}}(p) + d_{\mathrm{Ko}}(q) < C/p + C/q$ for any real p and q that are no less than 4. Taking (1.4) and (1.5) into account, one sees that (1.9) and (1.11) immediately yield better bounds:

Corollary 1.4. *For all p and q such that $4 \leqslant p < q \leqslant \infty$*

$$\tfrac{1}{2} d_{\mathrm{TV}}(p,q) = d_{\mathrm{Ko}}(p,q) < B(a,\tilde{x}_a) < C/p. \tag{1.12}$$

Graphs of the bounds $B(a,\tilde{x}_a)$ and C/p are shown in Figure 1, along with the corresponding graph of $d_{\mathrm{Ko}}(p)$. This is done for the values of $p \in [\frac{50}{29}, 30]$, even though the upper bound C/p on $d_{\mathrm{Ko}}(p)$ has been established only for $p \geqslant 4$.

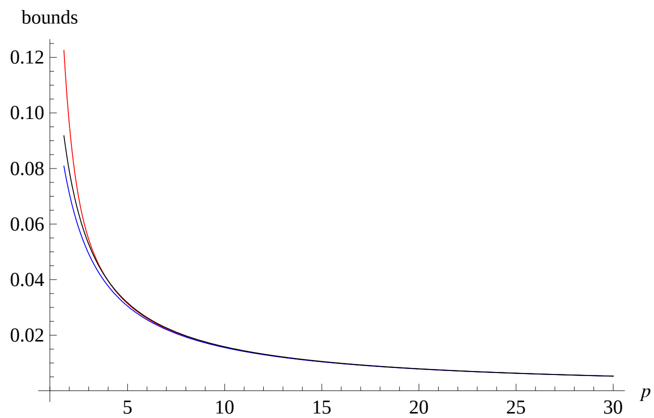

FIG 1. *Bounds $B(a,\tilde{x}_a)$ (red) and C/p (blue), compared with $d_{\mathrm{Ko}}(p)$ (black).*

The relative errors $\frac{B(a,\tilde{x}_a)}{d_{\mathrm{Ko}}(p)} - 1$ and $\frac{C/p}{d_{\mathrm{Ko}}(p)} - 1$ of the bounds in (1.9) and (1.6) are shown in Figure 2, for $p \in [1, 3.95]$ in the leftmost panel, for $p \in [3.95, 4.05]$ in the middle panel, and for $p \in [4.05, 30]$ in the rightmost one.

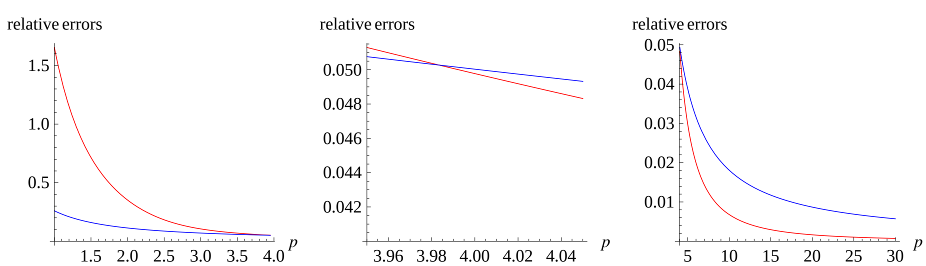

FIG 2. *Relative errors $\frac{B(a,\tilde{x}_a)}{d_{\mathrm{Ko}}(p)} - 1$ (red) and $\frac{C/p}{d_{\mathrm{Ko}}(p)} - 1$ (blue) of the bounds in (1.9) and (1.6).*

It appears that the bound C/p would be more accurate than $B(a, \tilde{x}_a)$ for $p \in [1, 3.98]$; remember, however, that the bound $B(a, \tilde{x}_a)$ was established only for $p \geqslant \frac{50}{29}$, from which the bound C/p was deduced only for $p \geqslant 4$. Anyway, the smaller values of $p > 0$ may be of lesser interest, since for such p the Student distribution is not very close to the standard normal one. On the other hand, for large p the bound $B(a, \tilde{x}_a)$ appears significantly more accurate (in terms of the relative errors) — but much more complicated — than the bound C/p. Yet, even for p as small as 4, the relative errors of the bounds C/p and $B(a, \tilde{x}_a)$ are both only about 5%, with the corresponding absolute errors less than 2×10^{-3}. For $p = 12$, the relative and absolute errors of the bound C/p are less than 1.5% and 2×10^{-4}, respectively, and the corresponding figures for the bound $B(a, \tilde{x}_a)$ are about 0.5% and 6×10^{-5}. Also, by (1.8), the relative error $\frac{C/p}{d_{\text{Ko}}(p)} - 1$ of the upper bound C/p goes to 0 as $p \to \infty$; in view of (1.11), the same holds for the upper bound $B(a, \tilde{x}_a)$. One may as well note that, if the distance $d_{\text{Ko}}(p)$ is considered as a kind of "initial" error — of the approximation of the Student distribution by the SND, then the relative error $\frac{C/p}{d_{\text{Ko}}(p)} - 1$ is a relative error "of the second order", in the sense that it is the relative error of the estimate C/p of the initial error $d_{\text{Ko}}(p)$; the same statement holds with $B(a, \tilde{x}_a)$ in place of C/p.

Figure 2 also suggests that the threshold value 4 in the condition $p \geqslant 4$ in Theorem 1.2 is very close to the best possible one for which the comparison (1.11) between the bounds in (1.6) and (1.9) is still valid.

In [7], an asymptotic expansion for the tail $1 - F_p(x)$ of the SD_p was obtained, which provides successive approximations (say $A_{p,j}(x)$) that are good for very large values of x, as illustrated in Figure 3 — for $p = 14$. The right panel

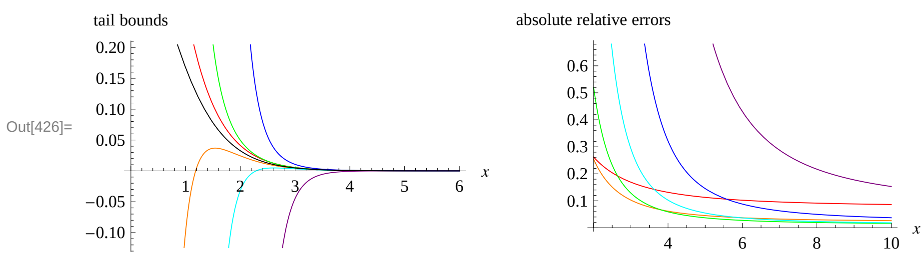

Fig 3. *Left panel: the successive approximations $A_{14,1}(x), \dots, A_{14,6}(x)$ of $1 - F_{14}(x)$ as in [7], colored red, orange, green, cyan, blue, purple, respectively; the graph of the tail $1 - F_{14}(x)$ is black. Right panel: the graphs of the corresponding absolute relative errors $|\frac{A_{14,j}(x)}{1-F_{14}(x)} - 1|$ ($j = 1, \dots, 6$).*

of Figure 3 suggests that each approximation $A_{p,j}(x)$ has its own "maximum competency" zone of values of x, for which it is the best, over all j's; it appears that this zone is a neighborhood of ∞, which gets narrower as j increases. Clearly, the bounds given in the present paper differ quite significantly in kind and purpose from those given in [7].

2. Proofs

The main idea of the proof of the inequality in (1.6) is to reduce it, through a number of steps, to systems of algebraic inequalities. Such systems, by a well-known result of Tarski [9, 1] (rooted in Sturm's theorem), can be solved in a completely algorithmic manner. Similar results hold for certain other systems which may also involve the logarithmic function (whose derivative is algebraic), the SND density function φ (whose logarithm is algebraic), and the SND distribution function Φ (whose derivative is φ). The bound $B(a, \tilde{x}_a)$ in Theorem 1.3 is such an expression. The Tarski algorithm is implemented in latter versions of Mathematica via `Reduce` and other related commands. For instance, a command of the form

```
Reduce[cond1 && cond2 && ..., {var1,var2,...,}, Reals]
```

returns a simplified equivalent of the given system (of equations and/or inequalities) `cond1`, `cond2`, ... over real variables `var1`, `var2`, ... However, the execution of such a command may take a very long time (and/or require too much computer memory) if the given system is more than a little complicated, as is e.g. the case with the system $B(a, \tilde{x}_a) < C/p$ & $a = 1/p$ & $p \geqslant 4$, which provides the way to deduce the bound in (1.6) from that in (1.9). Therefore, Mathematica will need some human guidance here. It appears that all such calculations done with the help of a computer are, at least, as reliable and rigorous as the same calculations done only by hand.

The main difficulty to overcome in this paper was to construct the upper bound $B(a, \tilde{x}_a)$ on $d_{\mathrm{Ko}}(p)$, which would be, on the one hand, accurate enough and, on the other hand, provide a traversable bridge from $d_{\mathrm{Ko}}(p)$ to the simple upper bound C/p, as indicated above. In turn, the bound $B(a, \tilde{x}_a)$ was obtained in several steps, described in detail in the statements of Lemmas 2.1–2.5, presented in Subsection 2.1.

The first step is to note that the difference $f_\infty(x) - f_p(x)$ between the densities of the SND and SD_p changes its sign exactly once, from $+$ to $-$, as x increases from 0 to ∞ (Lemma 2.1). A key observation here (essentially borrowed from [5]) is that, luckily, for the defined in (2.1) ratio $r_p(x)$ of the densities, the logarithmic partial derivarive $\frac{\partial}{\partial p} \ln r_p(x)$ increases in $x \in [0, 1]$ and decreases in $x \in [1, \infty)$ — with the same switch-point 1 for all $p > 0$. This implies that $r_p(x)$ decreases in $x \in [0, 1]$ from $r_p(0) < 1$ and then increases in $x \in [1, \infty)$ to ∞, so that the difference $F_\infty - F_p$ between the SND and SD_p distribution functions switches its monotonicity pattern just once — from increase to decrease — on the interval $[0, \infty)$, which provides a more manageable expression for the Kolmogorov distance $d_{\mathrm{Ko}}(p)$.

The next step concerns the difficulty that the expression (1.1) for $f_p(x)$ contains the so-called Wallis ratio $\Gamma(\frac{p+1}{2})/\Gamma(\frac{p}{2})$, which is not algebraic, and whose logarithm or derivative or logarithmic derivative is not algebraic either. To deal with this problem, we have just developed in [3] series of high-precision upper and lower algebraic bounds on the Wallis ratio; for the purposes of the present

paper, the first upper bound and the second lower bound in the corresponding series in [3] already suffice (Lemma 2.2). (A recent paper [2] provided other new upper and lower bounds on the Wallis ratio, improving on a number of preceding results. The series of bounds given in [3] (except a few first members of those series) are tighter than all the bounds in [2].) By using the mentioned results of [3], we obtain an upper bound, written as $H(a,x)/\sqrt{2\pi}$, on the difference $f_\infty(x) - f_p(x)$ between the densities of the SND and SD_p, which has an algebraic expression in place of the Wallis ratio (Lemma 2.3).

According to (2.3), $d_{\mathrm{Ko}}(p)$ equals a definite integral (in x) of the difference $f_\infty(x)-f_p(x)$; so, this integral can be bounded from above by the corresponding integral of the just mentioned upper bound $H(a,x)/\sqrt{2\pi}$. However, the latter integral is still problematic to estimate accurately enough. Toward that end, by some tweaking of the third-order Taylor polynomial in a for $H(a,x)$ near $a = 0$, we construct an upper bound $\tilde H_2(a,x)$ on $H(a,x)$, which is just the product of $\varphi(x)$ and a polynomial in a,x (Lemma 2.4). Thus, the bound $\tilde H_2(a,x)$ has certain nice properties (Lemma 2.5). Also, the relevant definite integral of $\tilde H_2(a,x)$ (corresponding to the mentioned one of $H(a,x)$) can be easily expressed in terms of the functions φ and Φ, thus finally resulting in the bound $B(a,\tilde x_a)$ in (1.9).

Inequality (1.11) (which, together with (1.9), yields the inequality in (1.6)) is provided by Lemma 2.6, whose proof is rather technical and relies on the Mathematica command `Reduce`, as described above. As for Proposition 1.1, it follows easily from Lemma 2.1 and the result of [5].

It appears that essentially the same method can be used to obtain even tighter upper (as well as lower) bounds on the distances $d_{\mathrm{Ko}}(p)$ and $d_{\mathrm{TV}}(p)$; toward such an end, one could use bounds in [3] on the Wallis ratio of higher orders of accuracy, as well as tweaked-Taylor polynomials for $H(a,x)$ of higher orders. The limitations on the attainable accuracy of such bounds on $d_{\mathrm{Ko}}(p)$ appear to be mainly set by the existing computational power; also, the proofs of the yet tighter bounds can be expected to be even more complicated.

In accordance with the above description of the scheme of proof, the current section is organized as follows. In Subsection 2.1, the mentioned lemmas are stated, thus presenting most of the main steps of proof. Next, in the same subsection, Proposition 1.1 and Theorems 1.3 and 1.2 are proved based on these lemmas. Finally, in Subsection 2.2 the lemmas stated in Subsection 2.1 (and requiring proof) are proved.

2.1. Statements of lemmas, and proofs of the main results

Introduce

$$r_{p,q}(x) := \frac{f_p(x)}{f_q(x)} \quad \text{and} \quad r_p(x) := r_{p,\infty}(x) = \frac{f_p(x)}{\varphi(x)}. \tag{2.1}$$

Lemma 2.1. *For each pair (p,q) such that $0 < p < q \leqslant \infty$*

(i) the ratio $r_{p,q}(x)$ decreases in $x \in [0,1]$ from $r_{p,q}(0) < 1$, and then increases in $x \in [1,\infty)$ to ∞; therefore,

(ii) *there is a unique point $x_{p,q} \in (0, \infty)$ (which is in fact greater than 1) such that*

$$\begin{aligned} &f_p(x) < f_q(x) \text{ for all } x \in [0, x_{p,q}), \\ &f_p(x_{p,q}) = f_q(x_{p,q}), \\ &f_p(x) > f_q(x) \text{ for all } x \in (x_{p,q}, \infty), \end{aligned} \tag{2.2}$$

and hence

$$d_{\mathrm{Ko}}(p, q) = F_q(x_{p,q}) - F_p(x_{p,q}). \tag{2.3}$$

For brevity, let

$$x_p := x_{p,\infty}. \tag{2.4}$$

Lemma 2.2. *For all real $p > 0$*

$$L_2(a) < r_p(0) < U_1(a), \tag{2.5}$$

where

$$L_2(a) := \frac{(1+2a)^{1/2}}{(1+a)^{7/8}(1+3a)^{1/8}} \quad \text{and} \quad U_1(a) := \frac{1}{(1+a)^{1/4}}.$$

This follows by the main result in [3]; the notations $r_p(0)$, $L_k(a)$, and $U_k(a)$ in the above Lemma 2.2 correspond to $r(p)$, $L_k(p)$, and $U_k(p)$ in [3].

The first inequality in (2.5), together with the definition (1.1), immediately yields

Lemma 2.3. *For all real $p > 0$ and all $x \in \mathbb{R}$*

$$f_\infty(x) - f_p(x) < \frac{H(a, x)}{\sqrt{2\pi}},$$

where

$$H(a, x) := e^{-x^2/2} - L_2(a)(1 + ax^2)^{-\frac{1+a}{2a}}.$$

By some tweaking of the third-order Taylor polynomial in a for $H(a, x)$ near $a = 0$, one obtains

$$\tilde{H}_2(a, x) := \frac{aP(a, x)}{384} e^{-x^2/2}, \tag{2.6}$$

where $P(a, x)$ is as in (1.10), so that $\tilde{H}_2(a, x)$ be an upper bound on $H(a, x)$:

Lemma 2.4. *For all $a \in (0, \frac{29}{50}]$ and $x \in (0, \frac{123}{50})$*

$$H(a, x) < \tilde{H}_2(a, x).$$

Lemma 2.5. *For each $a \in (0, 1)$, there is a unique real root $x > 0$ of the polynomial equation (1.10), so that $\tilde{x}_a$ is correctly defined in the statement of Theorem 1.3. Moreover,*

$$\begin{aligned} &\tilde{H}_2(a, x) > 0 \text{ for all } x \in (0, \tilde{x}_a), \\ &\tilde{H}_2(a, \tilde{x}_a) = 0, \\ &\tilde{H}_2(a, x) < 0 \text{ for all } x > \tilde{x}_a. \end{aligned}$$

Furthermore, $\tilde{x}_a$ is strictly and continuously increasing in $a \in (0, 1)$.

Lemma 2.6. *For all $p \geqslant 4$ inequality (1.11) holds.*

Now one is ready to prove Proposition 1.1 and Theorems 1.3 and 1.2, which will be done in this order.

Proof of Proposition 1.1. Take indeed any p and q such that $0 < p < q \leqslant \infty$. By Lemma 2.1 and the symmetry of the SD_p,

$$\begin{aligned} d_{\mathrm{TV}}(p,q) &= \int_0^{x_{p,q}} (f_q - f_p) + \int_{x_{p,q}}^{\infty} (f_p - f_q) \\ &= 2\int_0^{x_{p,q}} (f_q - f_p) = 2\big(F_q(x_{p,q}) - F_p(x_{p,q})\big) = 2d_{\mathrm{Ko}}(p,q), \end{aligned}$$

which proves part (i) of Proposition 1.1. Part (ii) of the proposition now follows by the second equality in (1.4) and the stochastic monotonicity result of [5], which implies that $F_p(x)$ is strictly increasing in $p \in (0,\infty]$ for each $x \in (0,\infty)$. □

Proof of Theorem 1.3. The equality in (1.9) immediately follows from Proposition 1.1. Take any $a \in (0, \frac{29}{50}]$ (corresponding to $p \geqslant \frac{50}{29}$). We claim that $x_p < \tilde{x}_a$, where x_p and $\tilde{x}_a$ are as in (2.4) and Lemma 2.5, respectively. Assume the contrary, that $x_p \geqslant \tilde{x}_a$. Note that $\tilde{H}_2(\frac{29}{50}, \frac{123}{50}) < 0$; so, by Lemma 2.5, $\tilde{x}_{29/50} < \frac{123}{50}$ and hence $\tilde{x}_a < \frac{123}{50}$ for all $a \in (0, \frac{29}{50}]$. Therefore, in view of Lemmas 2.3 and 2.4,

$$\sqrt{2\pi}\big(f_\infty(x) - f_p(x)\big) < H(a,x) < \tilde{H}_2(a,x) \tag{2.7}$$

for all $x \in (0, \tilde{x}_a]$ — still assuming that $a \in (0, \frac{29}{50}]$. On the other hand, by Lemma 2.1, $0 \leqslant f_\infty(x) - f_p(x)$ for all $x \in (0, x_p]$ and hence, by the assumption, for all $x \in (0, \tilde{x}_a]$. Now (2.7) implies $0 < \tilde{H}_2(a, \tilde{x}_a)$, which contradicts Lemma 2.5. Thus, indeed $x_p < \tilde{x}_a$. Recalling now (2.3) and using (2.7) and (again) Lemma 2.5, and also recalling (2.6), one has

$$\begin{aligned} d_{\mathrm{Ko}}(p) = F_\infty(x_p) - F_p(x_p) &= \int_0^{x_p} \big(f_\infty(x) - f_p(x)\big)\,\mathrm{d}x \\ &< \int_0^{x_p} \frac{\tilde{H}_2(x)}{\sqrt{2\pi}}\,\mathrm{d}x < \int_0^{\tilde{x}_a} \frac{\tilde{H}_2(x)}{\sqrt{2\pi}}\,\mathrm{d}x. \end{aligned}$$

It remains to verify that $\int_0^x \frac{\tilde{H}_2(u)}{\sqrt{2\pi}}\,\mathrm{d}u = B(a,x)$, which can be done either by hand or using Mathematica. The proof of Theorem 1.3 is now complete, modulo the lemmas. □

Proof of Theorem 1.2. The relations in (1.6) immediately follow by Theorem 1.3 and Lemma 2.6. It remains to verify (1.8). First here, use l'Hospital's rule to find that for all real x

$$\lim_{a\downarrow 0} \frac{f_{1/a}(x) - f_\infty(x)}{a} = \lim_{a\downarrow 0} \frac{\partial f_{1/a}(x)}{\partial a} = \lambda(x) := \frac{x^4 - 2x^2 - 1}{4}\,\varphi(x); \tag{2.8}$$

the second equality in (2.8) can be obtained either using the Mathematica commands `D` (for differentiation), `Simplify`, and `Limit` or otherwise.

Next, introduce

$$c_a := f_{1/a}(0) \quad \text{and} \quad g_a(x) = f_{1/a}(x)/c_a \tag{2.9}$$

for all real $a \geqslant 0$, assuming the convention $1/0 := \infty$, so that $f_{1/a}(x) = c_a g_a(x)$. Then for all real $a \geqslant 0$ and all real x

$$|f_{1/a}(x) - f_\infty(x)| \leqslant |c_a - c_0| g_a(x) + c_0 |g_a(x) - g_0(x)| \leqslant |c_a - c_0| + |g_a(x) - g_0(x)|, \tag{2.10}$$

since $g_a(x) \leqslant 1$ and $c_0 = 1/\sqrt{2\pi} < 1$. By (2.8) and (2.9), the ratio $\frac{|c_a - c_0|}{a}$ is continuous in $a > 0$ and converges to a finite limit $(\varphi(0)/4)$ as $a \downarrow 0$, and hence is bounded in $a \in (0, 1]$. Now note that

$$\Big|\frac{\partial g_a(x)}{\partial a}\Big| = (1 + ax^2)^{-(1+3a)/(2a)} \, |(Dg)(a, x)| \leqslant |(Dg)(a, x)|,$$

where

$$(Dg)(a, x) := \frac{\big(1 + ax^2\big) \ln\big(1 + ax^2\big) - a(1 + a)x^2}{2a^2}.$$

Using the Taylor expansion $\ln(1+u) = u - \theta u^2/2$ for $u > 0$ and some $\theta = \theta(u) \in (0, 1)$, one sees that $(Dg)(a, x)$ is a polynomial in a, x, θ and hence bounded in $(a, x) \in (0, 1] \times [0, \tilde{x}_0]$ — note that, in accordance with the definition of $\tilde{x}_a$ in Theorem 1.3,

$$\tilde{x}_0 = \sqrt{1 + \sqrt{2}} \in (0, \infty);$$

hence, $\Big|\frac{\partial g_a(x)}{\partial a}\Big|$ is bounded in $(a, x) \in (0, 1] \times [0, \tilde{x}_0]$ and, by the mean value theorem, so is $\frac{|g_a(x) - g_0(x)|}{a}$. Recalling also (2.10) and that the ratio $\frac{|c_a - c_0|}{a}$ is bounded in $a \in (0, 1]$, one concludes that the ratio $\frac{|f_{1/a}(x) - f_\infty(x)|}{a}$ is bounded in $(a, x) \in (0, 1] \times [0, \tilde{x}_0]$. So, by (2.8) and dominated convergence,

$$\begin{aligned} p \, d_{\mathrm{Ko}}(p) \geqslant p \big[F_\infty(\tilde{x}_0) - F_p(\tilde{x}_0)\big] &= -\int_0^{\tilde{x}_0} \frac{f_{1/a}(x) - f_\infty(x)}{a} \, \mathrm{d}x \\ &\underset{a \downarrow 0}{\longrightarrow} -\int_0^{\tilde{x}_0} \lambda(x) \, \mathrm{d}x = \frac{(\tilde{x}_0^3 + \tilde{x}_0) \, \varphi(\tilde{x}_0)}{4} = C, \end{aligned}$$

where $\lambda(x)$ is defined in (2.8). This, together with (1.6), implies (1.8). The proof of Theorem 1.2 is now complete, modulo the lemmas. □

2.2. Proofs of the lemmas

Proof of Lemma 2.1. Take indeed any p and q such that $0 < p < q \leqslant \infty$. A key observation here (borrowed from [5]) is that $r_{p,q}(x)$ decreases in $x \in [0, 1]$ and increases in $x \in [1, \infty)$. Moreover, by the lemma in [5], $f_p(0)$ increases in $p > 0$

and hence $r_{p,q}(0) < 1$. On the other hand, it is easy to see that $r_{p,q}(x) \to \infty$ as $x \to \infty$. This completes the proof of part (i) of Lemma 2.1, which in turn implies that there is a unique $x_{p,q} \in (0,\infty)$ such that $r_{p,q}(x) < 1$ for $x \in [0, x_{p,q})$, $r_{p,q}(x_{p,q}) = 1$, and $r_{p,q}(x) > 1$ for $x \in (x_{p,q}, \infty)$ (at that necessarily $x_{p,q} > 1$). In other words, one has the relations (2.2). Since $(F_q - F_p)' = f_q - f_p$, one now sees that $F_q(x) - F_p(x)$ increases in $x \in [0, x_{p,q}]$ from 0 to $F_q(x_{p,q}) - F_p(x_{p,q}) > 0$, and then decreases in $x \in [x_{p,q}, \infty)$ to 0. So, (2.3) follows by the symmetry of the Student and standard normal distributions. Thus, the lemma is completely proved. $\square$

Proof of Lemma 2.4. Indeed assume that $a \in (0, \frac{29}{50}]$ and $x \in (0, \frac{123}{50})$. Consider the difference

$$\tilde\delta(a) := \tilde\delta(a,x) := H(a,x) - \tilde H_2(a,x) = \frac{\tilde P(a,x)}{384 e^{x^2/2}} - L_2(a)(1+ax^2)^{-\frac{1+a}{2a}},$$

where

$$\tilde P(a,x) := 384 - aP(a,x).$$

We have to show that $\tilde\delta(a,x) < 0$. Obviously, the system of inequalities $\tilde P(a,x) \leqslant 0$, $0 < a \leqslant \frac{29}{50}$, and $0 < x < \frac{123}{50}$ is algebraic and thus, by the well-known result of Tarski [9] can be solved completely algorithmically. The Mathematica command `Reduce[tP<=0 && 29/50>=a>0 && 123/50>x>0]` outputs `False`, where `tP` stands for $\tilde P(a,x)$. This means that $\tilde P(a,x) > 0$ — for all $a \in (0, \frac{29}{50}]$ and $x \in (0, \frac{123}{50})$. So, $\tilde\delta(a,x)$ equals

$$\delta(a) := \delta(a,x) := \ln \frac{\tilde P(a,x)}{384 e^{x^2/2}} - \ln\Big(L_2(a)(1+ax^2)^{-\frac{1+a}{2a}}\Big)$$

in sign. Introduce

$$\begin{aligned}(D\delta)(a) &:= 4a^2\delta'(a) = \frac{4aQ(a,x)}{\tilde P(a,x)} - \frac{2(1+a)}{1+ax^2} - 2\ln\big(1+ax^2\big) \\ &\qquad + \frac{1}{6}\left(72a + \frac{21}{1+a} - \frac{6}{1+2a} + \frac{1}{1+3a} - 4\right), \\ (DD\delta)(a) &:= \frac{(D\delta)'(a)}{2a^3}\,(1+a)^2(1+2a)^2(1+3a)^2\big(1+ax^2\big)^2\,\tilde P(a,x)^2,\end{aligned}$$

where

$$Q(a,x) := a^3\big(20x^4 - 60x^2 - 33\big) - 96a\big(x^4 - 2x^2 - 1\big) - 768.$$

Note that $(DD\delta)(a)$ is a polynomial in a and x, of degree 11 in a and of degree 20 in x. The command `Reduce[DDde>=0 && 29/50>=a>0 && 123/50>x>0]` outputs `False`, where `DDde` stands for $(DD\delta)(a)$. This means that $(DD\delta)(a) < 0$ — for all $a \in (0, \frac{29}{50}]$ and $x \in (0, \frac{123}{50})$. On the other hand, one can check (using Mathematica or otherwise) that $(D\delta)(0+) = \delta(0+) = 0$. Thus, one concludes that indeed $\delta(a,x) < 0$ and hence $\tilde\delta(a,x) < 0$. $\square$

Proof of Lemma 2.5. Take indeed any $a \in (0,1)$. By (2.6), $\tilde{H}_2(a,x)$ equals $P(a,x)$ in sign. So, the first two sentences of Lemma 2.5 can be proved using the Mathematica command `Reduce[P>0 && 0<a<1 && x>0,x]`. That $\tilde{x}_a$ is strictly increasing in $a \in (0,1)$ now follows by the command `Reduce[PP[a,x]==0 ==PP[b,y] && 0<a<b<1 && 0<y<=x]`, which (takes about 15 seconds on a standard laptop and) outputs `False`; here, `PP[a,x]` stands for $P(a,x)$. Finally, the continuity of $\tilde{x}_a$ in a can be verified by the implicit function theorem; here, it is enough to check that $\frac{\partial P}{\partial x}(a,\tilde{x}_a) \neq 0$ for all $a \in (0,1)$, which can be done using the command `Reduce[P==0 && DPx==0 && 0<a<1 && x>0]` (with `DPx` standing for $\frac{\partial P}{\partial x}(a,x)$), which outputs `False`. □

Proof of Lemma 2.6. By Lemma 2.5, $a \mapsto \tilde{x}_a$ is a one-to-one map of $(0,\frac14]$ onto $(\tilde{x}_0,\tilde{x}_{1/4}]$. Let $(\tilde{x}_0,\tilde{x}_{1/4}] \ni x \mapsto a_x \in (0,\frac14]$ be the corresponding inverse map. So, it suffices to show that $B(a_x,x) < Ca_x$ for all $x \in (\tilde{x}_0,\tilde{x}_{1/4}]$. Assume indeed that $x \in (\tilde{x}_0,\tilde{x}_{1/4}]$ and consider the ratio

$$\rho(x) := \frac{B(a_x,x) - Ca_x}{a_x^3}. \tag{2.11}$$

Introduce also

$$\begin{aligned}
q_1(x) :=& 3 + 12x^2 + 30x^4 - 28x^6 + 3x^8, \\
q_2(x) :=& -783 - 2952x^2 - 1284x^4 + 2952x^6 - 234x^8 - 1608x^{10} \\
&+ 964x^{12} - 168x^{14} + 9x^{16}, \\
q_3(x) :=& 33 + 60x^2 - 20x^4.
\end{aligned}$$

The command `Reduce[q3 <= 0 && xxa[0] < x <= xxa[1/4]]` (with `q3` and `xxa[a]` standing for $q_3(x)$ and $\tilde{x}(a)$) outputs `False`, which shows that $q_3(x) > 0$. Now using the command `Reduce[P==0 && 0<a<=1/4 && xxa[0]<x<=xxa[1/4]]`, where P stands again for the polynomial $P(a,x)$ as in (1.10), one finds that

$$a_x = 2\,\frac{q_1(x) + \sqrt{q_2(x)}}{q_3(x)};$$

moreover, $a_x > 0$ and $q_3(x) > 0$ imply that $q_1(x) + \sqrt{q_2(x)} > 0$. So, in view of (2.11),

$$\begin{aligned}
&\rho'(x)\frac{e^{x^2/2}}{x}\,24\sqrt{\pi}\sqrt{q_2(x)}\big(q_1(x) + \sqrt{q_2(x)}\big)^3 \\
&\quad = \sqrt{2}\big(p_{00}(x) + p_{01}(x)\sqrt{q_2(x)}\,\big) + e^{x^2/2}\,C\sqrt{\pi}\big(p_{10}(x) + p_{11}(x)\sqrt{q_2(x)}\,\big),
\end{aligned}$$

where

$$\begin{aligned}
p_{00}(x) :=& 105705x - 1945539x^3 - 13305006x^5 - 26650971x^7 - 3174714x^9 \\
&+ 49512627x^{11} + 23388786x^{13} - 45078003x^{15} - 9879213x^{17} \\
&+ 26892909x^{19} - 5379786x^{21} - 8094383x^{23} + 6008972x^{25} \\
&- 1844301x^{27} + 296622x^{29} - 24435x^{31} + 810x^{33}, \\
p_{01}(x) :=& - 21789x - 259929x^3 - 492804x^5 + 366741x^7 + 967263x^9 \\
&- 120468x^{11} - 487080x^{13} + 188214x^{15} + 177266x^{17} - 151973x^{19} \\
&+ 43674x^{21} - 5625x^{23} + 270x^{25}, \\
p_{10}(x) :=& - 1368576 + 9287136x^2 + 67830048x^4 + 113324832x^6 - 52129440x^8 \\
&- 230541408x^{10} + 74263392x^{12} + 151161696x^{14} - 110996640x^{16} \\
&+ 30085440x^{18} - 3715200x^{20} + 172800x^{22}, \\
p_{11}(x) :=& 171072 + 1974240x^2 + 2340576x^4 - 4409568x^6 - 3045600x^8 \\
&+ 2911680x^{10} - 700800x^{12} + 57600x^{14}.
\end{aligned}$$

Executing now the command `Reduce[z<=0 && xxa[0]<x<=xxa[1/4]]` with `z` standing for $p_{10}(x) + p_{11}(x)\sqrt{q_2(x)}$, one sees that $p_{10}(x) + p_{11}(x)\sqrt{q_2(x)} > 0$, so that $\rho'(x)$ equals

$$\rho_1(x) := \frac{\sqrt{2}\big(p_{00}(x) + p_{01}(x)\sqrt{q_2(x)}\big)}{e^{x^2/2}\,\big(p_{10}(x) + p_{11}(x)\sqrt{q_2(x)}\big)} + C\sqrt{\pi}$$

in sign.

Next,

$$\begin{aligned}
\rho_2(x) :=& \rho_1'(x)\, e^{x^2/2}\,\big(p_{10}(x) + p_{11}(x)\sqrt{q_2(x)}\big)^2/\sqrt{2} \\
=& c_0(x) + c_1(x)\sqrt{q_2(x)} + c_2(x)q_2(x) + c_3(x)/\sqrt{q_2(x)},
\end{aligned}$$

where

$$\begin{aligned}
c_0(x) :=& p_{10}(x)p_{00}'(x) - p_{00}(x)p_{10}'(x) - xp_{00}(x)p_{10}(x), \\
c_1(x) :=& p_{11}(x)p_{00}'(x) - p_{00}(x)p_{11}'(x) - xp_{00}(x)p_{11}(x) + p_{10}(x)p_{01}'(x) \\
&- p_{01}(x)p_{10}'(x) - xp_{01}(x)p_{10}(x), \\
c_2(x) :=& p_{11}(x)p_{01}'(x) - p_{01}(x)p_{11}'(x) - xp_{01}(x)p_{11}(x), \\
c_3(x) :=& q_2'(x)\big(p_{01}(x)p_{10}(x) - p_{00}(x)p_{11}(x)\big)/2.
\end{aligned}$$

The command `Reduce[rho2[x]<=0 && xxa[0]<x<=xxa[1/4]]` outputs `False`. So, $\rho_2(x) > 0$ (for all $x \in (\tilde{x}_0, \tilde{x}_{1/4}]$) and hence $\rho_1(x)$ increases in such x. Moreover, $\rho_1(\tilde{x}_0) = 0$, which implies that $\rho_1 > 0$. That is, $\rho' > 0$ and ρ is increasing on the interval $(\tilde{x}_0, \tilde{x}_{1/4}]$, to $\rho(\tilde{x}_{1/4}) < 0$. Thus, $\rho < 0$ on $(\tilde{x}_0, \tilde{x}_{1/4}]$, which implies that indeed $B(a_x, x) < Ca_x$ for all $x \in (\tilde{x}_0, \tilde{x}_{1/4}]$. $\square$

References

[1] G. E. Collins. Quantifier elimination for real closed fields by cylindrical algebraic decomposition. In *Quantifier elimination and cylindrical algebraic decomposition (Linz, 1993)*, Texts Monogr. Symbol. Comput., pages 85–121. Springer, Vienna, 1998.

[2] C. Mortici. A new method for establishing and proving accurate bounds for the Wallis ratio. *Math. Inequal. Appl.*, 13(4):803–815, 2010.

[3] I. Pinelis. Geometrically convergent sequences of upper and lower bounds on the Wallis ratio and related expressions, preprint, http://arxiv.org/find/all/1/au:+pinelis/0/1/0/all/0/1.

[4] I. Pinelis. On the Berry–Esseen bound for the Student statistic, preprint, http://arxiv.org/find/all/1/au:+pinelis/0/1/0/all/0/1.

[5] I. Pinelis. Tail monotonicity properties of Student's family of distributions, preprint, http://arxiv.org/find/all/1/au:+pinelis/0/1/0/all/0/1.

[6] I. Pinelis and R. Molzon. Berry-Esséen bounds for general nonlinear statistics, with applications to Pearson's and non-central Student's and Hotelling's (preprint), arXiv:0906.0177v1 [math.ST].

[7] R. S. Pinkham and M. B. Wilk. Tail areas of the t-distribution from a Mills'-ratio-like expansion. *Ann. Math. Statist.*, 34:335–337, 1963.

[8] H. Scheffé. A useful convergence theorem for probability distributions. *Ann. Math. Statistics*, 18:434–438, 1947.

[9] A. Tarski. *A Decision Method for Elementary Algebra and Geometry.* RAND Corporation, Santa Monica, Calif., 1948.